Market Requirements Document

Feature Name: Resource Decription Framework (RDF) Support

Version: 2	Date Submitted: 12/24/2008	Completed By: Leon Guzenda
Version: 1	Date Submitted: 11/28/2007	Completed By: Leon Guzenda

Description of the Problem

Background

"The **Semantic Web** is an evolving extension of the World Wide Web in which web content can be expressed not only in natural language, but also in a format that can be read and used by software agents, thus permitting them to find, share and integrate information more easily.

The HTML language describes documents and the links between them. RDF, by contrast, describes arbitrary things such as people, meetings, or airplane parts. ." – Wikipedia

RDF Example (from Wikipedia)

In the English language statement 'New York has the postal abbreviation NY', 'New York' would be the subject, 'has the postal abbreviation' the predicate and 'NY' the object.

Encoded as an RDF triple, the subject and predicate would have to be resources named by URIs. The object could be a resource or literal element. For example, in the Notation_3 ("N-Triples") form of RDF, the statement might look like this:

<urn:states:New%20York> <http://purl.org/dc/terms/alternative> "NY" .

In this example:

- "urn:states:New%20York" is the URI for a resource that denotes the U.S. State New York;
- "http://purl.org/dc/terms/alternative" is the URI for a predicate;
- and "NY" is a literal string.

Note that the URIs chosen here are not standard, and don't need to be, as long as their meaning is known to whatever is reading them.

N-Triples is just one of several standard serialization formats for RDF. The triple above can also be equivalently represented in the standard RDF/XML format as:

```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:terms="http://purl.org/dc/terms/">
<rdf:Description rdf:about="urn:x-states:New%20York">
<terms:alternative>NY</terms:alternative>
</rdf:Description>
</rdf:RDF>
```

Problems

- 1. Objectivity/DB does not support the schema languages used by the various components of the semantic web stack, including RDF.
- 2. Objectivity/DB is not particularly efficient at storing and manipulating constructs that are primarily closely related text strings.
- 3. The internal Objectivity/DB schema representation is designed for relatively fixed structure languages, such as C++ and Java, rather than self-defining languages, such as Generalized Markup of Defined Objects (GDMO) and those used within the World Wide Web Consortium (W3C).

Impact

Customers wishing to store data from and for the Semantic Web will have to write their own metadata and data handling interfaces to Objectivity/DB, probably using Objectivity for Java. This is likely to be a considerable barrier to the adoption of Objectivity/DB in this rapidly evolving marketplace.

Description of the Requested Feature

This MRD is one of four MRDs covering Semantic Web environments. The others are:

- a) Semantic Web Support (High level index to the other three MRDs).
- b) XML Support for XML Schema and other XML standards, such as XPath and XQuery.
- c) Web Ontology Language (OWL).
- d) SPARQL Web query language.

The requested feature will provide:

- a) Efficient, scalable, high performance support for the description, storage, manipulation and querying of RDF schemas, triples and graphs. It should be able to handle trillions of triples.
- b) An RDF file loader and smart (resolves duplicates and disparities) reloader.
- c) Support for at least one popular RDF development framework, such as Jena
- d) Support for at least one popular inferencing framework, such as Jena 2.

Part of an existing feature or does it require another feature, if so, which one?

This feature will add a new Objectivity product – Objectivity/RDF.

How is this problem being solved now, and why isn't that acceptable?

Customers who need to add RDF support for the Semantic Web to existing products based on Objectivity/DB will have to build their own metadata and data handling tools. Prospects are more likely to look elsewhere for COTS solutions. There is a danger that some customers may also look for alternative solutions, either COTS or open source.

What languages must support this capability?

- Java
- C#
- Python (later).

Which platforms must be supported?

• Tier 1 at first and all platforms eventually.

Do any competitors already have this feature?

- Oracle and IBM have a growing number of RDF/OWL tools.
- <u>DATA-GRID</u> is developing a fully compliant, Internet enabled OWL database for delivery in Spring 2008. it will support RDF.
- Franz <u>AllegroGraph</u> is a high performance RDF Triple DBMS.

Customers who require this feature

- The Intelligence Community.
- The manufacturing, petrochemical, telecom and financial markets.
- Boeing has recently expressed interest in storing RDF and OWL in Objectivity/DB.

Revenue at risk, or which could be won

• <u>Analysts</u> predict that the global Semantic Web market could be worth \$75 Billion annually from 2010 onwards.

When is this required?

• Before 2009, if possible.

Additional Notes

- 1. We will also need:
- Marketing collateral and updated Technical Publications.
- New QA material.

2. We should consider a partnership arrangement with a commercial RDF/OWL tools vendor, such as <u>TopQuadrant</u>

3. If VArrays can be searched and indexed it may be possible to store an RDF arc as a Subject object with a VArray holding the Predicate and an association to an object representing the Object. RDF queries tend to start with a Subject or Object.